2010

Posture detection by kernel PCA-based manifold learning

Peng Cheng

University of Wollongong

Posture Detection by Kernel PCA-based Manifold Learning

A thesis submitted in fulfillment of the

requirements for the award of the degree

Master of Computer Science

from

UNIVERSITY OF WOLLONGONG

by

Peng Cheng

School of Computer Science and Software Engineering

September 2010

Dedicated to

Yuan-ying Cheng and Er-liang Zhang

Declaration

This is to certify that the work reported in this thesis was done by the author, unless specified otherwise, and that no part of it has been submitted in a thesis to any other university or similar institution.

Peng Cheng
September 29, 2010

Abstract

A Posture detection system aims to identify and localize any specific types of postures in images and video sequences. Unlike human or pedestrian detection where only one class of objects is required to be detected, posture detection is designed to detect multiple classes of postures. It remains a challenging problem because human bodies are complex and articulated with very diversified appearances. Posture detection often relies on a good generalization of the variations from large quantity of training examples that cover different situations. In this thesis, we devise a new posture detection framework that combines the histogram of gradient (HOG)-based feature with a novel manifold-based open-set classifier designed to achieve a better generalization. In this framework, each posture class is represented by a complex manifold that lies in the high-dimensional visual input space. The manifold is learned using Kernel PCA. Classification of a new observation is achieved by comparing it to each trained posture manifold. In addition, a new greedy Kernel PCA approximation algorithm is proposed to speed up the learning of the posture manifolds. The approximation algorithm seeks to remove the redundant training samples in the kernel space while best retaining the accuracy of kernel mapping, resulting in a new kernel PCA model that provides almost

identical learning and classification ability to the original kernel PCA with significantly lower computational cost. Both the detection framework and approximation algorithm were tested on 2D and 3D artificial datasets and real human and posture datasets. The results have shown that the approximation algorithm is effective and the proposed framework can provide accurate and efficient detection of different postures with a relatively small training set.

Acknowledgments

Great thanks to my supervisors, Associate Professor Wanqing Li and Professor Philip Ogunbona, for their guidance and demonstration of the required attitudes and principles of a researcher, and their constant encouragement and motivation. Also great thanks to everyone in the Advanced Multimedia Research Lab: Ce Zhan, Nguyan Duc Thanh, Alister Coordiner, Shenbo Guo, Jun Hu, Zhenqiang Jiang and Zhou Sun for sharing their ideas and offering constructive discussions. This thesis would have not not been completed without their help.

Contents

List of Tables

List of Figures

0.1 Glossary

Adaboost : Adaptive boosting, a boosting algorithm for ensemble learning.

auROC : Area under Receiver's Operating Characteristic curve, a scalar equal to the integration of an ROC curve from 0 to 1, used to measure the performance of classifiers.

DET : Decision Error Trade-off curve, a curve showing the missing rate versus the false positive per window (FPPW).

DR : Detection Rate, a scalar denoting the percentage of successful detection per window.

DT : Distance Transform, a 2D pixel map showing the distance from pixels to a given contour.

DoG : Derivative of Gaussian.

FPPW : False Positive per Window, a scalar denoting the percentage of false detection per window.

GMM : Gaussian Mixture Model, a statistic model for probabilistic estimation of a multivariate distribution.

HMM : Hidden Markov Model, a random process model primarily used in modeling temporary events.

HOG : Histogram of Gradient, a visual feature mainly used in human detection and body representation.

ICA : Independent Component Analysis: a toolset consisting of several matrix decomposition and factorization techniques that aim to isolate statistically mutually independent bases from a set of data.

ISM : Implicit Shape Model, an object detection framework by synthesizing local detection results through spatial voting.

Isomap : Isometric mapping, a nonlinear manifold embedding technique to reconstruct a mapping space that preserves the graph-distance.

KPCA :Kernel Principal Component Analysis, a generalized nonlinear manifold learning framework that performs the principal component analysis in a kernelized feature space.

KNN :K-Nearest Neighbors, a simple example-based classifier that labels each unknown datum to the majority of its K-Nearest Neighbors.

LBP : Local Binary Pattern, a visual descriptor that is invariant to illumination.

LLE : Locally Linear Embedding: a nonlinear manifold embedding technique to reconstruct a mapping space that minimizes the change of distances between adjacent data.

LPP : Locally Preserving Projection, a linear dimensionality reduction technique that aims to find a linear mapping to a lower dimension space that minimizes the same objective function with Laplacian eigenmap.

MDS : Multidimensional Scaling, a toolset that aims to reconstruct a equivalent dataset in an explicit Euclidean space from a similarity or dissimilarity matrix.

MHI : Motion History Image, a 2D greyscale image characterising motion information of a binary video.

MRF : Markov Random Field, a 2D statistic graphic model used to model spatially correlated random variables.

NMF : Non-negative Matrix Factorization, a matrix decomposition and factorization technique that aims to isolate non-negative bases from a dataset.

PCA : Principal Component Analysis.

RANSAC : Random Sample Consensus, a fast model estimation meta-algorithm.

RBF : Radial Basis Function, a bivariate function in the form of $f(||x-y||^2)$.

RMI : Recurrent Motion Image, a 2D greyscale image characterising recurrent motion information of a binary video.

ROC : Receiver's Operating Characteristic curve, a curve showing the detection rate versus the false positive detection.

ROI : Region of Interest.

SIFT : Scale Invariant feature transform.

SVM : Support Vector Machine.

0.2 Notations

$E(.)$: expectation of a variable or a set of random vectors.

$cov(.)$: covariance matrix of a variable or a set of random vectors.

$p(X)$: marginal probability or likelihood of a random variavle X.

$p(X|Y)$: conditional probability or likelihood of X given Y.

X : the column matrix of a training dataset.

x_i : the i^th vector in the training dataset.

c_i : the class of the i^th example.

$\phi(.)$: the implicit non-linear mapping that maps [.] into an infinite-dimensional feature space.

Φ : the implicit column matrix of $\phi(x_i)$.

$k(.,.)$: the positive semidefinitive bivariate kernel function that defines $\phi(.)$ by its inner-product.

K : the inner-product matrix of $\phi(x_i)$ where $K_{ij} = k(x_i, x_j)$.

H : the constant matrix for centralizing data in the feature space. $H = I_n - 1_n$, I_n is an $n \times n$ identity matrix and 1_n denotes a $n \times n$ matrix in which each element takes the value of $1/n$.

$\hat{\Phi}$: the implicit column matrix of $\phi(x_i)$ centered at zero mean.

$\hat{\phi}(x_i)$: the i^th column vector of $\hat{\Phi}$.

$\hat{K}$: the inner-product matrix of $\hat{\Phi}$, $\hat{K} = HKH$.

λ_i : the i^th eigenvalue.

D : the diagonal matrix with each diagonal element $D_{ii} = \lambda_i$.

P : the implicit column matrix of eigenvectors of $cov(\hat{\Phi})$ (also known as principal components).

A : the column matrix of eigenvectors of $\hat{K}$.

z : an arbitrary datum sample in the test dataset to be classified.

c : the ground truth of z.

$y(.)$: the function that maps z into the subspace of principal components in the feature space.

H_A : the abbreviation for HA, it is the most important matrix in defining the KPCA model and the projection $y(.)$.

$w(.)$: polynomial part of $y(.)$ defined by $w(.) = (H_A)^T k(X,.)$, where $k(X,.) = [k(x_1,.), k(x_2,.) \ldots k(x_n,.)]^T$

b : the constant part of $y(.)$ defined by $b = (H_A)^T K 1_n$

U : the multidimensional scaling result of K.

$\tilde{X}$: the column matrix of a subset of the training dataset selected by a kernel approximation algorithm.

$\tilde{P}$: shortened P for faster KPCA mapping $y(.)$.

$\tilde{H}_A$: shortened H_A for faster KPCA mapping $y(.)$.

$W_{\tilde{X}}$: the w-mapping of the subset $\tilde{X}$ into the KPCA mapping space (the column space of the original principal components P).

Q **and** R : QR decomposition of $D_n^{\frac{1}{2}} W_{\tilde{X}}$.

V : the set of $D_n^{\frac{1}{2}} W_X$ orthonormalized with $W_{\tilde{X}}$.